ChinaXiv合作期刊

个体视角下的网络舆情传递链路预测分析*

魏　静[1]　朱恒民[1]　宋瑞晓[2]　蒋世兵[3]
[1](南京邮电大学管理学院　南京 210023)
[2](南京航空航天大学经济与管理学院　南京 210016)
[3](布鲁克大学管理学院　圣凯瑟琳斯 L2S 3A1)

摘要:【目的】通过网络舆情传播中网民个体的“从众效应”和“阈值效应”的传递特点分析，对舆情传递的链路结构进行预测。【方法】采集真实网络舆情传递在线数据，利用舆情传递个体的节点属性及舆情传递网络结构的已知信息，使用链路预测方法，分别针对网络舆情无标度(BA)网络模拟数据和真实舆情传递网络数据，预测已有节点间即将产生的舆情传递连接。【结果】通过网络数据仿真和 BBS 真实数据链路预测分析发现，在众多的链路预测相似性指标中，局部路径指标(LP)算法得出的链路预测结论正确率最高，说明 LP 算法适合此类舆情传递网络的链路预测分析。【局限】仅限于对已有链路预测相似性指标的应用，没有对传统链路预测相似性指标进行相应的改进。【结论】从数据角度提供一个有效的预测舆情发展趋势的分析方法，以期为网络舆情控制提供相关理论支持。

关键词: 链路预测　舆情个体　无标度网络　BBS 网络
分类号: G35　C93

chinaXiv:201711.01259v1

1　引　言

中国拥有世界上最多的网民和最大的访问量，有效分析网络舆情，对于政府可以更好地解读公众意见，对于媒体可以突破传统的单向信息生产模式，实现信息增值。因此，政府、媒体、社会公众都越来越重视网络舆情分析，充分发挥舆情信息价值。然而，要充分分析网络舆情，仅仅依靠传统的舆情分析方法“被动接受网络舆情数据”是远远不够的。由于网络在线数据采集、清洗的滞后性，在线数据的分析通常不能充分反映瞬息万变的舆情发展状况。

网络舆情的监控主要是指发现对网络舆情出现、发展和消亡具有重要影响的因素，并连续不间断地进行动态监测度量、采集信息，根据预警体系的内容运用综合分析技术对当前网络舆情做出评价分析，进而及时地做出等级预报活动[1]。如兰月新等[2]构建三个网络舆情维度的舆情危机评估指标体系，使得舆情预警研究由趋势性预警转变为防御性预警。其他比较常用的网络舆情分析方法主要有：网络调查法、基于统计规则的模式识别方法和基于内容挖掘的主题监测方法[3]。

虽然，现有文献的研究内容已经开始涉及“防御性预警”，但是研究的基础仍然是“过去的舆情数据”，即使采用“正在发生的舆情数据”，只是对舆情现有状况的一个评估，其分析结果仍然摆脱不了滞后性。在当今的数据环境下，舆情演变瞬时爆发，既然不能推演出舆情“即时分析结果”，那么是不是可以尝试“领先一步，预测需求”？因此，本研究尝试针对已有的舆情数据预测将来的舆情走向。

常见的网络行为预测模型中，线性模型包括：泊松模型[4]、马尔科夫模型[5]、自回归模型 AR[6]、自回

通讯作者：魏静，ORCID：0000-0002-0730-4547，E-mail：mirror820909@163.com。

*本文系国家自然科学基金项目“互联网舆情演化中群体行为协同演进模型研究”(项目编号:71271120)、教育部人文社会科学基金项目“基于互联网舆情传递阈限的群体事件在线监控研究”(项目编号:13YJC630178)和江苏省社会科学基金项目“大数据时代网络舆情引发群体事件的在线监控研究”(项目编号:14TQC002)的研究成果之一。

ChinaXiv合作期刊

chinaXiv:201711.01259v1

归平均模型 ARMA[6]和时间序列预测模型[7]等，这些预测模型算法相对简单，数据拟合比较方便，但是对于非线性、长相关、自相似的舆情传递个体数据，不能完全适用；非线性预测模型中最典型的神经网络模型[8]，具有良好的自学习、自适应及泛化能力的优点，但是神经网络模型比较适用于指数增长的预测，对具备明显的无标度及择优增长特性的互联网舆情传递数据，预测效果不够理想。

本研究选择采用复杂网络理论中的链路预测分析方法，进行舆情传递预测，具体原因如下：

(1) 复杂网络理论可以准确描述互联网舆情非线性、长相关以及自相似的传递状态；

(2) 链路预测算法对舆情传递过程中的无标度、择优增长特性更加关注，算法突出了舆情传递过程中主要的网络增长特点及传播特性；

(3) 本研究的目的要求输出舆情传递的走势，链路预测算法与此相吻合，更加直接地反映出舆情传递下一时段传递链接的增长情况以及未来链接产生的可能性。希望通过链路预测算法探测舆情传递网络结构的演化，实现预测舆情传递的目的，以期为网络舆情的监控和治理提供数据支持。

2 舆情传递个体的信息传递关系解析

2.1 舆情传递的"从众效应"

从实质上来讲，舆情传递过程可以看作个人意见的表达。而个人意见的表达是一个社会心理表达过程，其发生的根本原因是为了防止个体因孤立而受到社会惩罚。个人在表明自己观点之际，首先会对周围环境中存在的意见进行观察和检测。当个体发现自己属于"多数意见"或者"优势意见"持有者时，他们便拥有了"安全感"，同时倾向于积极地表明自己的观点。相反，当个体发现自己属于"少数意见"或者"意见劣势"状态时，个体便会屈服于"环境压力"，选择沉默状态，或者放弃表达自己意见的权利[9]。

学者们通常把舆情传递过程中存在的类似现象称为"从众效应"[10]。个体倾向于与群体内其他成员建立和维持一致性，刘锦德[11]认为，网络舆情传播过程是以大部分人对事实不了解为前提的，因此从众行为是网络舆情得以扩散的一个重要原因。"从众效应"的产生并不是仅仅由个体对舆情信息的有效性和有利性做出判断，而是来自于个体因未获得信息而丢失个体优势的恐慌，本文将这种"恐慌"称之为"从众压力"。如果舆情传递个体认为吸收舆情信息会产生个体损失，或者认为吸收舆情信息的个体收益不够高而放弃舆情信息，此时"从众压力"便会产生，如图 1 所示：

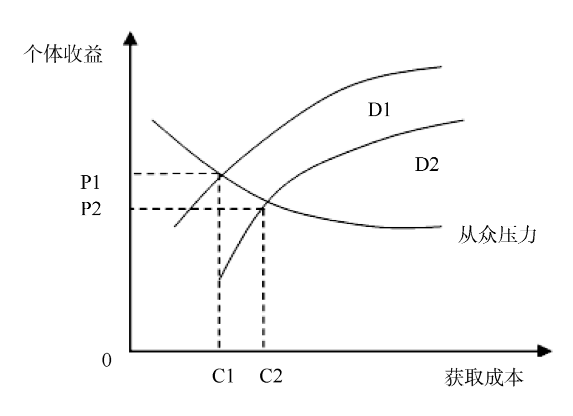


图 1 "从众压力"产生图示

最初认为舆情信息无效的个体，会在"从众压力"的作用下，选择接受舆情信息。这些屈服于"从众压力"的个体行为导致吸收舆情信息的个体数量增多，从而强化了"从众压力"。最终，"从众行为"引发了一个循环驱动：舆情信息接收个体数量的增多引发了"从众压力"，"从众压力"反过来导致舆情信息接收者的增多。这种循环驱动在没有外力的情况下会一直持续，直到认为舆情信息无效的所有个体屈服于"从众压力"，而最终接受舆情信息。或者，另外一种情况，"从众行为"的循环驱动止于舆情个体对"从众压力"的"不屈服"反应。

2.2 舆情传递的"阈值效应"

在"从众行为"的定义中，一部分舆情传递个体会首先接收舆情信息，另一部分舆情传递个体选择追随其后。"从众行为"促使组织产生良性的反馈回路，首先接受信息的个体带来了"从众压力"，"从众压力"促使更多的舆情个体接受舆情信息。在这一过程中，可以看到，为什么个体对舆情信息的吸收有先有后？为什么不同舆情传递个体存在不同的舆情信息接受倾向？

其根本原因来自于舆情传递个体传播"阈值"的存在。舆情是社会公众所参与的群体意识活动，舆情传递个体的性格、舆情传递个体间的关系及舆情传递个体文化背景等因素存在差异，本文将不同舆情传递个体对某一特定舆情信息的接收程度定义为舆情信息接

ChinaXiv合作期刊

chinaXiv:201711.01259v1

收“阈值”。在舆情传递过程中，舆情传递个体的吸收阈值被逐渐增大的“从众压力”超越，最终舆情传递个体屈服于“从众压力”而变成舆情信息接收者，从而导致舆情传递组织内舆情信息接收个体的比例进一步增大，如图 2 所示。本文将阈值的这一作用过程定义为“阈值效应”。

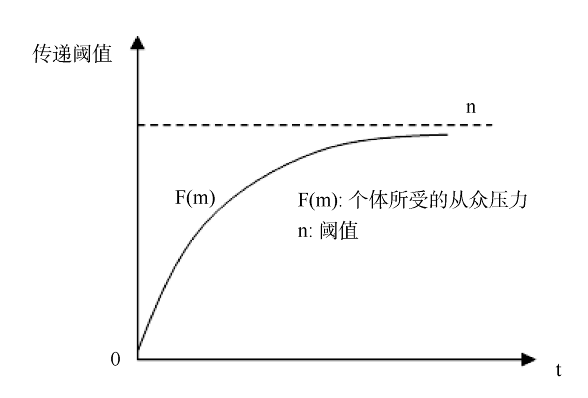


图 2　舆情传递个体传播“阈值”的作用过程

对于阈值较高的舆情传递个体，可以承受较高的“从众压力”，只有“从众压力”足够大时，舆情传递个体才会屈服并接收舆情信息。对于阈值较低的舆情传递个体，所能承受的“从众压力”较低，在较低的“从众压力”下便会接受舆情信息。对于持有“0 阈值”的舆情传递个体，不能承受任何的“从众压力”，对舆情信息没有抵触倾向，可以很轻易地接受任何舆情信息，从而反向增大舆情组织内部的“从众压力”。

如此，循环往复，信息接收者的比率增大，助长了“从众压力”的增大，同时引发了更多的舆情信息接收者，如图 3 所示:

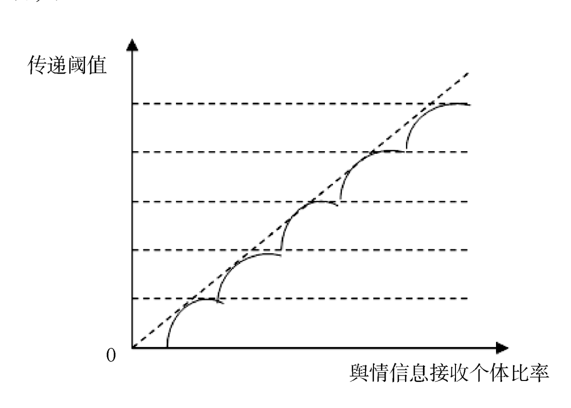


图 3　传递阈值循环作用图示

当舆情组织内部的“从众压力”强度不足以达到舆情信息未接收者的阈值，并且不能成功引起舆情传递个体信息接受行为时，循环终止。“阈值效应”可以解释：为什么“从众效应”没有使所有的舆情传递个体均变为信息接收个体。实际上，存在另外一种情况，在舆情传递的任何阶段，当所有舆情信息未接收节点的传递阈值均超越“从众压力”，“从众效应”也会终止，即所有个体均变为舆情信息接收个体。

3　链路预测分析

3.1　相关研究

链路预测本质上是通过已知网络节点的节点属性和网络结构特征等信息，估计两个尚未产生连接的节点之间产生连边的可能性，或者是已经实际存在，但是目前仍然没有被发现的连边。

目前，链路预测分析方法被应用到生物、互联网、社会网络等各个领域。比如在生物领域预测蛋白质网络结构中的关键蛋白，代替耗费大量成本的实验室操作[12]；在社会网络领域，通过挖掘隐藏边信息来挖掘恐怖分子，阻止恐怖事件的发生[13]。另外，很多学者致力于链路预测方法本身的改进。Liben-Nowell 等[14]针对社会网络中的合作网络特征重新定义了符合社会网络特征的相似性指标。Murata 等[15]还将共有邻居指数(CN)拓展到加权网络中，Leskovec 等[16]研究了在线社交网络上的正负关系的链路预测。

舆情传递领域的链路预测是将链路预测分析方法应用于社会网络领域，通过已有的舆情传递网络拓扑结构，预测未来某一时刻的舆情传递拓扑结构，从而得知舆情传递的走向和路径。另外，本研究将各种相似性指标算法的正确率进行比较，试图找出适合特定舆情传递网络的相似性指标计算方法。

3.2　链路预测

舆情信息的流动是表达舆情传递个体间网络结构的“从众效应”和表达舆情传递个体的主观情绪的“阈值效应”共同作用结果。因此有学者指出，舆情传递个体的行为规则不仅决定于个人意志，而且取决于舆情传递个体所处的位置[17]。网络结构能够影响舆情传递个体间的从众关系，因为在舆情传递个体的“视线”之外的舆情接收者是没有办法对潜在舆情传递个体产生影响的。同时，在舆情接收者的可影响范围内，其释放的“影响”无处不渗透着舆情传递个体的阈值底线。

另外，结构上对等的个体(即存在相似的舆情信息传递邻居)面临着相似的"从众环境"和"阈值环境"，因此如此相似的环境下，潜在舆情接收者也会表现出相似的舆情信息反应。那么，能不能通过已知的舆情传递个体间的"从众效应"关系和舆情传递个体的"阈值效应"属性状态预测舆情传递个体间的未来舆情信息传播状态？即预测舆情传递个体间进行舆情传递的可能性。对这一问题的回答，是本文的研究重点，本文借用链路预测的分析方法。

链路预测是复杂网络理论研究中的一个重要分支，旨在利用已知的网络拓扑结构或者节点属性信息预测网络中节点对之间缺失或者可能产生的链接，其实质是探索和预测网络信息传递结构的演化。比如对于生物网络中隐含未知链接的揭示是需要耗费高额实验成本的，但是如果可以预测，而非盲目地检测所有链接，并以此指导实验，就可以节约相当一部分实验开销。对于不断演化的舆情传递网络而言，根据已知的舆情传递个体间的传递和影响关系，预测尚未连接的舆情传递个体间的扩散和传递状态，从而避免对海量信息的盲目筛选，实现对舆情传递的预期和预测。由此揭示舆情传递个体间的传递和演化行为，制定相应的舆情治理策略，改变了舆情治理只能事后操作的传统模式，转为舆情事前监督。

4 链路预测仿真

4.1 无标度(BA)网络模型介绍

在复杂网络理论的各种网络模型中，随机网络(ER)和小世界网络(WS)的共同特征是网络的度分布可近似表示为泊松分布，因此可以被称为均匀网络。但是，大量实证研究结果表明，舆情传递所借助的在线社交网络度分布具有更突出的幂率分布特性——大部分的节点度比较小，少数节点拥有较大的节点度[18-19]。

幂率分布是BA无标度网络的一个重要特征，BA无标度网络的各个节点之间的连接具有严重的不均匀分布特性，网络中只有少部分的节点拥有极其多的连接，而大多数节点拥有少量连接。BA无标度网络的这一度分布特点与在线社交网络度分布的特点相吻合，因此其通常被认定为模拟真实在线网络结构的理想网络[20-21]。

BA无标度网络的构造算法可以分为两步：

(1) 增长：从一个具有 m_0 个节点的网络开始，每次引入一个新的节点，并且连接到 m 个已经存在的节点上，取 $m_0 \geq m$ 。

(2) 优先连接：一个新节点与一个已存在的节点 i 相连接的概率 p，节点 i 的度 k_i 和节点度的和 $\sum_j k_j$ 之间满足如下关系：

$$p = \frac{k_i}{\sum_j k_j} \quad (1)$$

4.2 无标度舆情信息传递网络模型构建

由于"阈值效应"存在，舆情不断扩散的同时会引起舆情传递网络不断增长，舆情传递过程中"从众效应"引起的优先连接的特性，最重要的是舆情传递网络具备无标度网络的幂率性特征[22]，因此选取无标度网络作为舆情传递链路预测的仿真网络。

结合网络舆情的传递特点，基于 BA 无标度网络模型构造网络模型算法[22]：

(1) 始化网络：设网络初始节点数为 $m_0 = 20$，为每个网络节点随机设置不同的初始阈值 TR_0 。由于舆情信息随着时间的推移，其时效性会存在一定程度的减退，故设定网络的衰减参数 $d = 0.2(0 \leq d < 1)$。

(2) 入新节点：在网络中加入 l 个新节点，新节点的舆情传递阈值仍然随机产生，新节点将与网络中原有节点产生 $m(m \leq m_0)$ 条连边。在连接过程中，由于从众效应的存在，原有节点 i 的优先连接选取概率 π_i 不仅与节点 i 的传递阈值有关，而且与网络原有节点度的增加系数 $\alpha \in (0,1)$ 有关，取 $\alpha = 0.1$，即：

$$\pi_i = \frac{TR(i) + \alpha}{\sum_j TR(j)} \quad (2)$$

(3) 新网络节点的阈值：随着舆情环境的改变，个体的舆情传递阈值也会发生变动。

$$\begin{pmatrix} TR(1) \\ \vdots \\ TR(l) \end{pmatrix} = (1-d)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + d\frac{A_{il}}{m}\begin{pmatrix} TR(1) \\ \vdots \\ TR(l) \end{pmatrix} \quad (3)$$

其中，$A_{i \times l}$ 为网络的链接矩阵。

(4) 复节点加入，直至节点增加满足要求 m_t=300。

"优先连接"是 BA 网络构建的一个特性，本研究用 BA 网络的优先连接特性表达舆情传递的"从众效应"。网络环境中个体"从众效应"越强，其选择建立连

chinaXiv:201711.01259v1

ChinaXiv合作期刊

chinaXiv:201711.01259v1

接的节点数量越多。为了表达舆情传递过程中“从众效应”的大小，在构造 BA 网络的时候，新加入节点与网络原有节点建立连接时分别选取 $m=1,2,3,4$，建立 4 个不同的 BA 网络。在经过 BA 网络的“增长”和“优先连接”过程之后，所生成舆情传递 BA 无标度网络的具体网络相关统计指标如表 1 所示：

表 1 网络连边数

网络	Net1	Net2	Net3	Net4
边数	450	730	941	1 140
节点数	300	300	300	300

4.3 链路预测实现

参考文献[23]中的链路预测步骤，结合本文的研究内容，分别对 4 个 BA 网络模型进行链路预测：

(1) 按照随机抽样方法，以 80%和 20%的比例，将网络连边数据随机分成训练集 E^T 和测试集 E^P 两部分。

(2) 设网络中节点个数为 V=300，已有连边组成的集合为 E_{tra}。则网络中未连接的节点连边为 $V \times V - E_{tra}$。

(3) 根据不同的相似性指标计算方法，遍历所有节点对 $<x,y>$，计算相似性指标分数 $score(x,y)$。相似性指标分数值是所有连边的相似性指标分数值，但是需要关注目前实际上不存在的连边的相似性指标值。

(4) 将计算出的相似性指标分数大小进行排序，按照从大到小的顺序填到排序表格中，相似性指标的分数越大，表明节点对间出现连边的几率越大。

(5) 设测试集 E^P 中连边集合为 E_{pro}，网络中实际连边的个数为 $n=E_{pro}$。选取相似性指标分数列表中的前 n 对节点建立连边，这些连边为预测出的存在可能性最大的连边，设为集合 E_{pre}。

(6) 检验预测出连边的正确性，设检验正确率的指标为 P：

$$P = \frac{|E_{pre} \cap E_{pro}|}{n} \times 100\% \quad (4)$$

经过程序测算，得到不同相似性指标测算方法得出的测算值，相关指标解释如表 2 所示。其中，对于网络中的节点 x，定义 $\Gamma(x)$ 是 x 的邻居集合，s_{xy} 代表点 x 和 y 的共同邻居个数；k_x, k_y 分别为节点 x 和 y 的度；A 为网络的邻接矩阵，α为可调参数，I 为单位矩阵；Katz 指标中α需小于邻接矩阵最大特征值的倒数。

表 2 链路预测相似性指标

名称	定义
共同邻居(CN)	$s_{xy}=\|\Gamma(x)\cap\Gamma(y)\|$
Salton 指标	$s_{xy}=\frac{\|\Gamma(x)\cap\Gamma(y)\|}{\sqrt{k_x k_y}}$
Jaccard 指标	$s_{xy}=\frac{\|\Gamma(x)\cap\Gamma(y)\|}{\|\Gamma(x)\cup\Gamma(y)\|}$
Sorensen 指标	$s_{xy}=\frac{2\times\|\Gamma(x)\cap\Gamma(y)\|}{k_x+k_y}$
大度节点有利指标(HPI)	$s_{xy}=\frac{\|\Gamma(x)\cap\Gamma(y)\|}{\min\{k_x,k_y\}}$
大度节点不利指标(HDI)	$s_{xy}=\frac{\|\Gamma(x)\cap\Gamma(y)\|}{\max\{k_x,k_y\}}$
LHN-I 指标	$s_{xy}=\frac{\|\Gamma(x)\cap\Gamma(y)\|}{k_x k_y}$
优先连接指标(PA)	$s_{xy}=k_x k_y$
Katz 指标	$s=(I-\alpha\times A)^{-1}-I$
局部路径指标(LP)	$s=A^2+\alpha A^3$

计算其对应的正确率检验指标 P，结果如表 3 所示：

表 3 预测正确率(%)

指标	Net1	Net2	Net3	Net4
CN	18.45	10.61	10.31	7.52
Salton	0.06	1.06	4.23	6.40
Jaccard	0.06	1.12	4.30	6.43
Sorensen	0.06	1.12	4.30	6.43
HPI	0.04	0.05	0.32	0.20
HDI	0.02	0.04	0.26	0.23
LHN-I	0.01	0.02	0.27	0.41
PA	20.70	**14.81**	**11.62**	**10.52**
Katz	14.32	11.31	9.72	9.35
LP	**21.60**	12.61	10.47	9.65

链路预测结果表明，针对本次仿真实验的 BA 网路数据，PA 相似性指标的预测结果较为准确。这与 BA 网络的特征有密切关系，BA 网络的构建是通过引入节点、优先连接的步骤实现的。PA 相似性指标的计算思路也是侧重于优先连接的考虑，因此针对 BA 网络模型数据，PA 相似性指标的预测结果较为准确。

另外，LP 相似性指标的预测结果相比其他相似性指标的预测方法，也呈现出比较好的准确性。这是因为 LP 是局部路径相似性指标，而 BA 网络的典型特征

chinaXiv:201711.01259v1

是度分布具备幂率分布特性，局部群体特征相对其他类型网络更为明显。因此，针对本研究的实验数据，LP相似性指标预测方法也比较适用。

5 舆情传递个体信息传递的链路预测验证

5.1 舆情传递个体数据采集描述

除了 Blog、Wiki 等形式以外，BBS 是网络舆情传递系统的重要组成部分之一。随着 Web 页面的使用普及，BBS 逐渐拥有了庞大的用户群，同时满足网络舆论的产生及传递有效性的要求。

本文所使用的舆情现实数据来源于江苏省南京市某企业在线员工论坛①，该 BBS 论坛有 300 多个用户，工作日在线率较高(75%以上)，讨论版面齐全，为本文的研究提供了良好的原始数据源。BBS 数据集在非监督的情况下通过网络爬虫程序获得，不需要用户配合，数据真实可靠。所获取的原始数据由 2014 年 7 月到 2015 年 6 月共计约 134 369 条的 BBS 发帖和回帖记录组成。利用数据库技术和数据挖掘技术挖掘有关信息，对所采集的数据建立 BBS 在线网络。该 BBS 在线网络是以发帖或回帖个体为节点、回帖为边的网络。一个帖子的发帖人和回帖人，可能同时与另一个帖子或者其他几个帖子的发帖人和回帖人产生交叉，因此不同帖子的发帖人和回帖人共同组成了网络系统，整个网络的度分布如图 4 所示。网络的度分布具备明显的长尾现象，因此具备无标度网络的幂率分布特性。

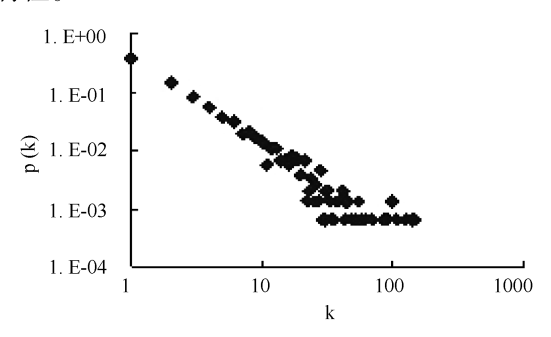


图 4 舆情传递网络度分布

由于构建出的 BBS 网络系统节点及连边规模比较大，经过初步数据清洗之后，选择论坛的两个板块作为分析的主体。两个板块命名为“advice-4”和“advice-5”，其网络基本统计指标如表 4 所示：

表 4 advice-4 和 advice-5 的网络基本属性特征值

统计指数	advice-4	advice-5
节点数	33	18
连接数	103	65
密度	0.0985	0.2157
互惠性	1	1
传递性	0.1615	0.3542
聚集系数	0.303	0.505
平均距离	2.746	2.190
网络直径	6	4
连通性	1	1
网络效率	0.9597	0.8824

5.2 现实舆情传递网络的链路预测

针对选定的论坛版块 advice-4 和 advice-5 进行链路预测分析：

(1) 针对清洗后的数据，划分训练集和测试集。将 2014 年 7 月–2014 年 12 月的数据作为训练集，将 2015 年 1 月–2015 年 6 月的数据作为测试集。

(2) 得到 2014 年 12 月某时间点，advice-4 和 advice-5 中的节点数分别为 G_4、G_5 和连边数 E_4、E_5。两个板块中不存在的连边数分别为 $(G_4 \times G_4 - E_4)$、$(G_5 \times G_5 - E_5)$。

(3) 根据相似性指标计算方法，遍历所有节点对 $<x, y>$，计算节点对的相似性指标分数 score(x, y)，并依据指标数值大小排序。

(4) 虽然企业员工相对固定，但是偶有新员工加入，导致论坛节点数量稍有增加。而本文链路预测分析主要关注“原有节点”间连边数量的增加。故采用文献[24]的处理方法，选取 2014 年 12 月某时刻之前的连边组成的集合为 E_{former}。实际新产生的连边为：$E^*_{former} = E_{former} \cap (G \times G)$。则新增连边的数目为：$n = |E^*_{former}|$。

(5) 选取相似性指标列表里面的前 n 对节点建立的连边为预测出的连边，设此预测连边的集合为 E_{pre}。

(6) 用各种方法计算相似值的链路预测正确率 P：

①http://168.8.4.18/bbs/index.php.

ChinaXiv合作期刊

chinaXiv:201711.01259v1

$$P = \frac{\left|E^{*}_{former} \cap E_{pre}\right|}{n} \times 100\% \quad (5)$$

5.3　链路预测结果分析

经过上述步骤的链路预测，得到预测正确率结果如表 5 所示。

链路分析结果表明 LP 算法的正确率最高，LP 相似指标计算的主要思路是考虑主要邻居及次级邻居的共同影响。将抓取到的所有 BBS 在线数据进行可视化处理，如图 5 所示。

可以清晰地看到网络舆情扩散呈现很明显的分层传递态势。虽然链路预测只针对此舆情传递网络的部分数据，但是仍然存在相似的分层传播现象。因此，结合 LP 相似性指标的层级影响思路，得出的预测结果比较准确。同时，LP 指标链路预测结果的准确性，也从另一方面肯定了仿真模型建立的基本思路——“从众效应”和“阈值效应”对舆情传递网络理论解析的准确性。

表 5　预测正确率结果比较

指标	advice-4 正确率(%)	advice-5 正确率(%)	advice-4 预测条数	advice-5 预测条数
CN	20.10	15.61	32	22
Salton	8.21	9.27	4	2
Jaccard	10.35	11.60	3	3
Sorensen	10.35	11.60	3	3
HPI	7.62	6.01	1	1
HDI	8.12	7.42	2	2
LHN-I	1.09	1.11	0	0
PA	18.39	14.03	25	18
Katz	19.07	13.21	31	20
LP	26.72	18.39	40	24

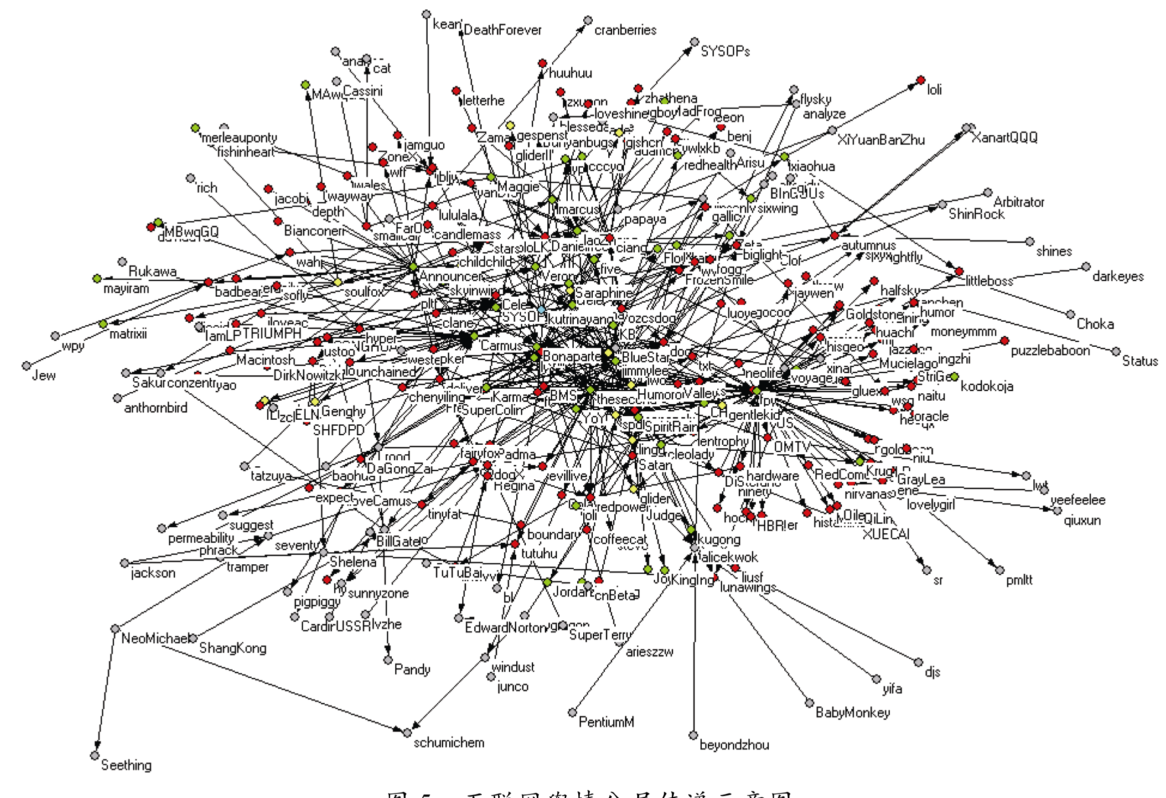


图 5　互联网舆情分层传递示意图

(■: 核心节点 ■: 一级节点 ■: 二级节点 ■: 三级节点)

另外，CN 相似性指标的算法准确率也相对较高，CN 算法的主要判断标准是共同邻居的数量。通过规则对等性的测算，对 CN 算法的计算结果进行验证。

在 advice-4 网络中，规则对等性测算结果如图 6 和图 7 所示。

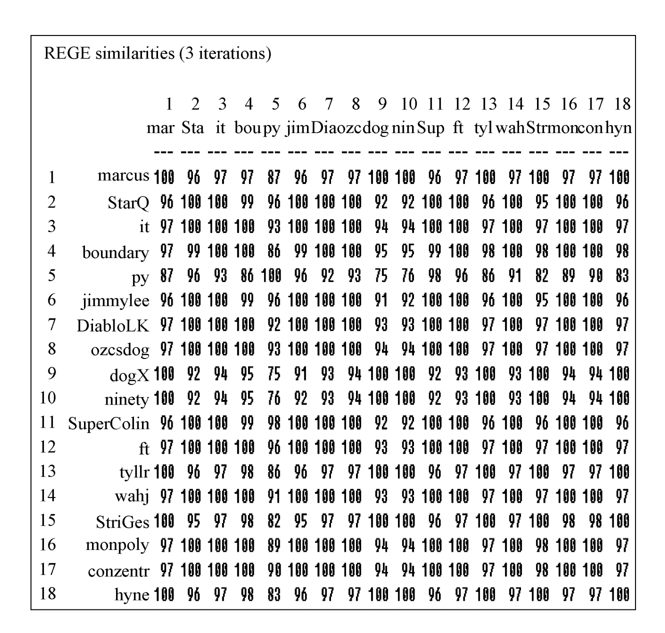
(1) 5、28、18、22、25、29 等节点在 100%的相似水平上具有规则对等性;

chinaXiv:201711.01259v1

RESG similarities (3 iterations)

		1 roo	2 plt	3 ble	4 HEL	5 sma	6 sev	7 shi	8 lul	9 liu	10 cle	11 Ann	12 OMT	13 wsg	14 qoi	15 iwo
1	rood	100	99	100	100	99	92	99	99	100	99	99	99	100	99	99
2	pltf	99	100	99	99	100	99	100	100	99	100	92	100	99	100	93
3	blessedareye	100	99	100	100	99	92	99	99	100	99	99	99	100	99	99
4	HELP	100	99	100	100	99	94	99	99	100	99	91	99	100	99	92
5	smallcai	99	100	99	99	100	99	100	100	99	100	93	100	99	100	94
6	seventy	92	99	92	94	99	100	99	99	91	99	80	99	93	99	82
7	shingo	99	100	99	99	100	99	100	100	99	100	92	100	99	100	93
8	lululala	99	100	99	99	100	99	100	100	99	100	92	100	99	100	93
9	liusf	100	99	100	100	99	91	99	99	100	99	99	99	100	99	99
10	cleolady	99	100	99	99	100	99	100	100	99	100	90	100	99	100	92
11	Announcer	99	92	99	91	93	80	92	92	99	90	100	92	94	94	100
12	OMTV	99	100	99	99	100	99	100	100	99	100	92	100	99	100	93
13	wsg	100	99	100	100	99	93	99	99	100	99	94	99	100	99	95
14	qoiooo	99	100	99	99	100	99	100	100	99	100	94	100	99	100	95
15	iwo	99	93	99	92	94	82	93	93	99	92	100	93	95	95	100
16	Saraphine	100	99	100	100	99	92	99	99	100	99	99	99	100	99	99
17	chenyiling	100	99	100	100	99	91	99	99	100	99	99	99	100	99	99
18	CONGHUA	99	100	99	99	100	99	100	100	99	100	93	100	99	100	94
19	ustoo	100	99	100	100	99	92	99	99	100	99	99	99	100	99	99
20	spdf	100	99	100	100	99	91	99	99	100	99	99	99	100	99	99
21	SpiritRain	100	99	100	100	99	91	99	99	100	99	99	99	100	99	99
22	MPC	99	100	99	99	100	99	100	100	99	100	93	100	99	100	94
23	lvzhe	99	100	99	99	100	99	100	100	99	100	90	100	99	100	92
24	Judge	99	100	99	99	100	99	100	100	99	100	90	100	99	100	92
25	BMS	99	100	99	99	100	99	100	100	99	100	89	100	99	100	90
26	tramper	96	99	95	97	99	100	99	99	95	99	86	99	96	99	88
27	NeoMichael	99	100	99	99	100	99	100	100	99	100	92	100	99	100	93
28	pker	99	100	99	99	100	99	100	100	99	100	93	100	99	100	94
29	skygragon	99	100	99	99	100	99	100	100	99	100	92	100	99	100	93
30	KB	99	100	99	99	100	99	100	100	99	100	90	100	99	100	92
31	adamchang	99	100	99	99	100	99	100	100	99	100	90	100	99	100	92

图 6 advice-4 规则对等性相似矩阵(部分数据)



REGE similarities (3 iterations)

		1 mar	2 Sta	3 it	4 bou	5 py	6 jim	7 Dia	8 ozc	9 dog	10 nin	11 Sup	12 ft	13 tyl	14 wah	15 Str	16 mon	17 con	18 hyn
1	marcus	100	96	97	97	87	96	97	97	100	100	96	97	100	97	100	97	97	100
2	StarQ	96	100	100	99	96	100	100	100	92	92	100	100	96	100	95	100	100	96
3	it	97	100	100	100	93	100	100	100	94	94	100	100	97	100	97	100	100	97
4	boundary	97	99	100	100	86	99	100	100	95	95	99	100	98	100	98	100	100	98
5	py	87	96	93	86	100	96	92	93	75	76	98	96	86	91	82	89	90	83
6	jimmylee	96	100	100	99	96	100	100	100	91	92	100	100	96	100	95	100	100	96
7	DiabloLK	97	100	100	100	92	100	100	100	93	93	100	100	97	100	97	100	100	97
8	ozcsdog	97	100	100	100	93	100	100	100	94	94	100	100	97	100	97	100	100	97
9	dogX	100	92	94	95	75	91	93	94	100	100	92	93	100	93	100	94	94	100
10	ninety	100	92	94	95	76	92	93	94	100	100	92	93	100	93	100	94	94	100
11	SuperColin	96	100	100	99	98	100	100	100	92	92	100	100	96	100	96	100	100	96
12	ft	97	100	100	100	96	100	100	100	93	93	100	100	97	100	97	100	100	97
13	tyllr	100	96	97	98	86	96	97	97	100	100	96	97	100	97	100	97	97	100
14	wahj	97	100	100	100	91	100	100	100	93	93	100	100	97	100	97	100	100	97
15	StriGes	100	95	97	98	82	95	97	97	100	100	96	97	100	97	100	98	98	100
16	monpoly	97	100	100	100	89	100	100	100	94	94	100	100	97	100	98	100	100	97
17	conzentr	97	100	100	100	90	100	100	100	94	94	100	100	97	100	98	100	100	97
18	hyne	100	96	97	98	83	96	97	97	100	100	96	97	100	97	100	97	97	100

图 8 advice-5 规则对等性相似矩阵

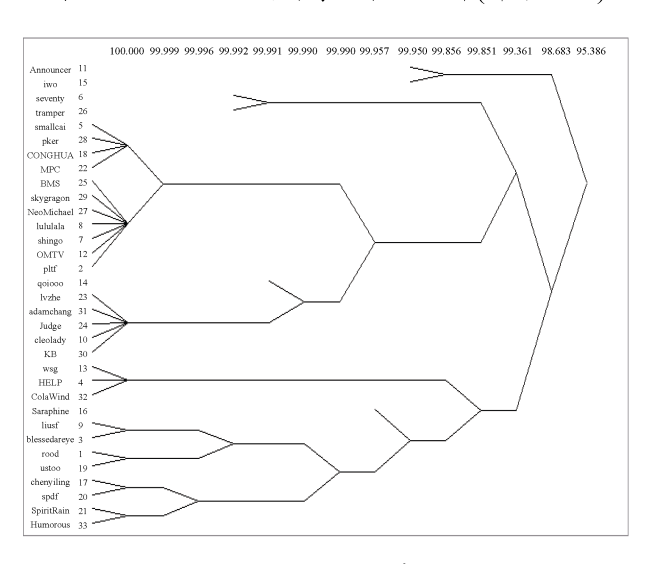


图 7 advice-4 规则对等性聚类树

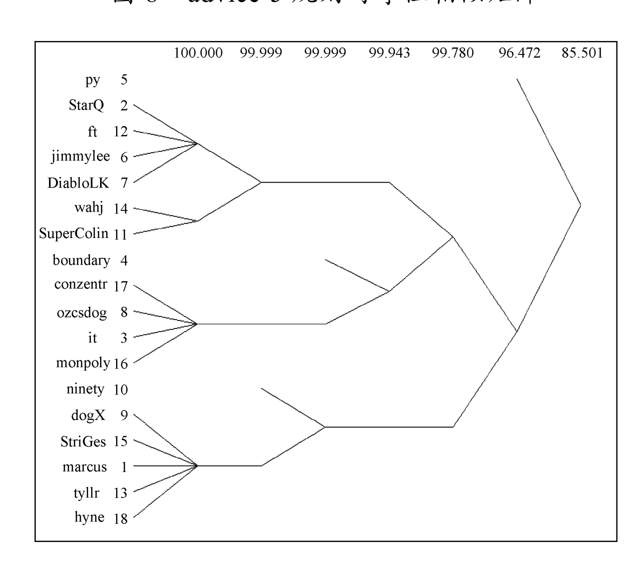


图 9 advice-5 规则对等性聚类树

(2) 节点 6 和节点 26 间在 99.991%的水平上具备规则对等性。这与 CN 算法计算出的(5, 29)、(28, 18)、(5, 22)、(6, 26)等节点(只列举部分预测连接数据)对间的相似性指标值最高的结果基本相吻合。

在 advice-5 网络中, 规则对等性测算结果如图 8 和图 9 所示。

(1) 2、12、6、7 等节点在 100%的相似水平上具有规则对等性;

(2) 节点 5 在 86%的相似水平上与节点 10 具有规则对等性, 在 95%的相似水平上与节点 10 具有规则相似性。这一结果与 CN 算法得到的(12, 7)、(6, 2)、(5, 9)等节点对间的相似性指标值较高的结果也基本吻合。

真实网络的实证研究结果表明, 舆情传播节点之间能建立联系, 受他们共同邻居的影响[25]。同样, 在网络舆情传递过程中, 不相干的舆情信息传递个体间能否建立直接联系, 很大程度上取决于他们是否具有共同邻居, 具有共同邻居的数量越大, 建立直接联系的概率就越高。

因此可以得出结论: 针对此舆情传递 BBS 网络, CN 算法的预测结果最精确, 可以使用 CN 算法对此舆情传递网络做链路预测分析。

ChinaXiv合作期刊

6 结 语

通过理论分析发现，网络舆情传递个体间的"从众效应"和"阈值效应"是舆情得以传递的重要驱动力。以其为理论基础，对舆情传递 BA 模拟网络数据和真实舆情传递网络的 BBS 在线数据进行链路预测分析。在对众多的相似性指标计算结果比较以后发现，LP 算法的计算结果准确率最高，CN 算法的准确率其次。并且使用可视化及对等性分析等手段，对链路预测的计算结果进行了相关验证。

本研究也存在一些不足，在对仿真数据和真实数据进行链路预测时，选取的目前学术领域经常采用的一些相似性计算指标，而没有对其进行相应的改进，使指标更能贴合舆情传递网络的特点，这也是后续研究的方向所在。

参考文献：

[1] 吴绍忠，李淑华.互联网络舆情预警机制研究[J]. 中国人民公安大学学报：自然科学版，2008，14(3)：38-42. (Wu Shaozhong, Li Shuhua. Research on Early Warning Mechanism for Online Public Opinion [J]. Journal of the People's Public Security University: Natural Science Edition, 2008, 14(3): 38-42.)

[2] 兰月新，董希琳，陈成鑫. 地方政府应对网络舆情能力评估和危机预警研究[J]. 现代情报，2012，32(5)：8-12. (Lan Yuexin, Dong Xilin, Chen Chengxin. Research on Local Governments to Deal with Network of Public Opinion Crisis Warning[J]. Journal of Modern Information, 2012, 32(5): 8-12.)

[3] 陈忆金，曹树金，陈少驰，等. 网络舆情信息监测研究进展[J]. 图书情报知识，2011(6)：41-49. (Chen Yijin, Cao Shujin, Chen Shaochi, et al. Survey on Online Public Opinion Information Monitoring [J]. Document, Information & Knowledge, 2011(6): 41-49.)

[4] 梁娜. 基于 ANN-泊松模型的预测研究[J]. 咸宁学院学报，2008, 28(6): 22-24. (Liang Na. Research on Prediction Based on ANN-Poisson Model [J]. Hournal of Xianning University, 2008, 28(6): 22-24.)

[5] 尹清波，张汝波，李雪耀，等.基于线性预测与马尔可夫模型的入侵检测技术研究[J].计算机学报，2005，28(5)：900-907. (Yin Qingbo, Zhang Rubo, Li Xueyao, et al. Research on Technology of Intrusion Detection Based on Linear Prediction and Markov Model [J]. Chinese Journal of Computers, 2005, 28(5): 900-907.)

[6] 范玉妹，玄婧.ARMA 算法在 GDP 预测中的应用[J]. 江南大学学报：自然科学版，2010，9(6)：736-740. (Fan Yumei, Xuan Jing. Application of ARMA Model to GDP Prediction [J]. Journal of Southern Yangtze University: Natural Science Edition, 2010, 9(6): 736-740.)

[7] 杨云霞.时间序列预测模型及其应用[J]. 太原师范学院学报：自然科学版，2005，4(4)：4-7. (Yang Yunxia. Prediction Model and Application of Temporal Sequence [J]. Journal of Taiyuan Normal University: Natural Science Edition, 2005, 4(4): 4-7.)

[8] 杜华英. 基于 PSO 算法的 BP 神经网络研究[J]. 现代计算机，2009(2)：22-27. (Du Huaying. Research on BP Neural Network Based on PSO Algorithm [J]. Modern Computer, 2009(2): 22-27.)

[9] Taylor S E, Peplou L A, Sears D O.社会心理学[M]. 第 10 版. 谢晓非，谢冬梅，张怡玲，等译. 北京：北京大学出版社，2004：325-326. (Taylor S E, Peplou L A, Sears D O. Psychology [M]. The 10th Edition. Translated by Xie Xiaofei, Xie Dongmei, Zhang Yiling, et al. Beijing: Beijing University Press, 2004: 325-326.)

[10] Abrahamson E, Rosenkopf L. Institutional and Competitive Bandwagons: Using Mathematical Modeling as a Tool to Explore Innovation Diffusion[J]. Academy of Management Review, 1993, 18(3): 487-517.

[11] 刘锦德.基于不完全信息演化博弈模型的网络舆情传播羊群行为[J].国防科技大学学报，2013，35(5)：96-101. (Liu Jinde. Herb Behavior in Dissemination of Public on the Internet Based on Evolutionarey Game Model with Incomplete Information [J]. Journal of National University of Defense Technology, 2013, 35(5): 96-101.)

[12] Maslov S, Sneppen K. Specificity and Stability in Topology of Protein Networks [J]. Science, 2002, 296(5569): 910-913.

[13] Lam L. Histophysics: A New Discipline [J]. Modern Physics Letters B, 2002, 16(30): 1163-1176.

[14] Liben-Nowell D, Kleinberg J. The Link Prediction Problem for Social Networks [J]. Journal of the American Society for Information Science and Technology, 2007, 58(7): 1019-1031.

[15] Murata T, Moriyasu S. Link Prediction of Social Networks Based on Weighted Proximity Measures [C]. In: Proceedings of the 2011 IEEE/WIC/ACM International Conference on Web Intelligence. IEEE, 2007: 85-88.

[16] Leskovec J, Huttenlocher D, Kleinberg J. Predicting Positive and Negative Links in Online Social Networks [C]. In: Proceedings of the 19th International Conference on World Wide Web. 2010: 641-650.

[17] 吴少华，崔鑫，胡勇. 基于 SNA 参数增量的网络舆论调控

chinaXiv:201711.01259v1

ChinaXiv合作期刊

时机研究[J]. 网络安全技术与应用, 2014(6): 167-168. (Wu Shaohua, Cui Xin, Hu Yong. A Study on the Timing to Guide Intemet Public Opinion Based on SNA [J]. Net Security Technologies and Application, 2014(6): 167-168.)

[18] Ahn Y Y, Han S, Kwak H, et al. Analysis of Topological Characteristics of Huge Online Social Networking Services [C]. In: Proceedings of the 16th International Conference on World Wide Web. 2007: 835-844.

[19] Fu F, Chen X, Liu L, et al. Social Dilemmas in an Online Social Network: The Structure and Evolution of Cooperation [J]. Physics Letters A, 2007, 371(1): 58-64.

[20] Mislove A, Marcon M, Gummadi K P, et al. Measurement and Analysis of Online Social Networks [C]. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. 2007: 29-42.

[21] Kumar R, Novak J, Tomkins A. Structure and Evolution of Online Social Networks [A]. // Social Networking[M]. Springer International Publishing, 2010: 337-357.

[22] 魏静, 朱恒民, 宋瑞晓, 等.在线知识转移网络的演化规律实证分析[J]. 管理评论, 2014, 26(12): 38-44. (Wei Jing, Zhu Hengmin, Song Ruixiao, et al. An Empirical Research on Evolution Rules of Online Knowledge Transfer Network [J]. Management Review, 2014, 26(12): 38-44.)

[23] 吕琳媛, 周涛. 链路预测[M]. 北京: 高等教育出版社, 2013. (Lv Linyuan, Zhou Tao. Link Prediction [M]. Beijing: Higher Education Press, 2013.)

[24] 王林, 商超. 无标度网络中的链路预测问题研究[J]. 计算机工程, 2012, 38(2): 67-70. (Wang Lin, Shang Chao. Research on Link Prediction Problem in Scale-free Network [J]. Computer Engineering, 2012, 38(2): 67-70.)

[25] 田占伟. 基于复杂网络的微博信息传播研究[D]. 哈尔滨: 哈尔滨工业大学, 2012. (Tian Zhanwei. Research on Microblog Inforamtion Transmission Based on Complex Network [D]. Harbin: Harbin Institute of Technology, 2012.)

作者贡献声明:

魏静, 蒋世兵: 提出研究思路, 设计研究方案, 进行实验;
朱恒民: 采集、清洗、分析数据;
宋瑞晓: 论文修改及最终版本修订;
魏静, 朱恒民, 宋瑞晓, 蒋世兵: 论文起草。

收稿日期: 2015-07-17
收修改稿日期: 2015-09-24

Link Prediction Analysis of Internet Public Opinion Transfer from the Individual Perspective

Wei Jing[1] Zhu Hengmin[1] Song Ruixiao[2] Jiang Shibing[3]
[1] (School of Management, Nanjing University of Posts & Telecommunications, Nanjing 210023, China)
[2](College of Economic and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)
[3](Department of Management, Brock University, St. Catharines L2S 3A1, Canada)

Abstract: [**Objective**] This paper establishes the BA network model of public opinion transfer process, regarding "Bandwagon Effect" and "Threshold Effect" as a starting point and according to the special inspection of public opinion. [**Methods**] At the same time, collect the real online data of public opinion transfer network. This paper uses the link prediction method to predict the unknown links of public opinion nodes which will appear in the forthcoming transfer process of both simulation BA network data and real public opinion data. [**Resualts**] The analysis results show that among many similarity indices algorithms LP link prediction algorithm can get the best prediction. It means that LP link prediction algorithm is suitable for the link prediction in such public opinion delivery network. [**Limitations**] There is no improvement of link predict similarity index. [**Conclutions**] From the point of data view, this paper proposes an effective prediction method of public opinion trends analysis to provide the theoretical support for the network of public opinion control.

Keywords: Link prediction Individual of public opinion transfer BA network BBS network

chinaXiv:201711.01259v1